DOCUMENT RESUME

ED 037 710

AL 002 335

AUTHOR Robinson, Peter.
TITLE Oral Expression Tests.
PUB DATE Jun 69
NOTE 17p.; Based on a paper presented at the Canadian Linguistics Association Congress, Toronto, June 10-13, 1969

EDRS PRICE EDRS Price MF-$0.25 HC-$0.95
DESCRIPTORS Communication (Thought Transfer), *Conversational Language Courses, English (Second Language), Evaluation Methods, Language Proficiency, *Language Tests, *Linguistic Competence, Linguistic Performance, Listening Comprehension, Objective Tests, *Oral Expression, Reading Comprehension, Second Languages, Speech Skills

ABSTRACT

The discussion centers around the amount and kind of participation that should be asked of a student taking reading and oral comprehension tests. For each test there are four levels of participation; zero, limited, extended, and complete. The test designer should create a situation which will encourage the subject to give maximum expression, because his language in the test is a representative sample of what he can produce in a real situation. This also enables the corrector to give a more accurate evaluation of the student's language ability. The most effective oral test includes an interview and a discussion section to permit the most extensive use of the language. Various ways of evaluating linguistic structures are suggested, with emphasis on organization of test goals. (FB)

ED037710

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

# O R A L  E X P R E S S I O N  T E S T S [1]

by

Peter ROBINSON

Head TESOL Programme
Etudes anglaises
Université Laval

A L 0 0 2 3 3 5

---

1- Based on a paper given at the Canadian Linguistics Association Congress, held at Toronto from June 10 to June 13, 1969.

Oral expression tests are on the whole neglected as it is impossible to treat the subject's answers in an objective form. Instead of arriving at a certain numeral through counting a series of ticks on answer sheets, the corrector is confronted with a large number of phrases whose grammatical, lexical, phonetic and socio-cultural content has to be evaluated. The evaluation has to be simple and complete; standardised according to formal criteria so that the correction is rapid and consistent, and can easily be explained to other correctors who achieve the same rapidity and consistency, whatever their knowledge of linguistics may be.

The term objective is a misleading term. It implies that the results of an objective test are in themselves objective, when in fact they are no more reliable than those of a subjective test. What is it then that makes an objective grammar test more meaningful and reliable than a subjective translation test? What does 53% mean in objective test? Interpretation of the figure 53% is a subjective affair, as is the evaluation of the grammar content of a translation.

The basic differences between the objective and subjective test are not of objectivity versus subjectivity, of scientific method versus intuitive appreciation. The first difference is one of form: the way in which the question is asked, the way the answer is given and the way the answer is corrected. The second difference is that the objective test assesses the subject's language competence on an extremely limited number of items. The third difference is that the objective test evaluates the subject's language competence passively: passive in that the subject makes little or no use of the language; passive in that the subject's knowledge of the language is evaluated rather than his capacity to use the language. It is therefore clear that the objective test by limiting the subject's freedom of expression tends

to become a tabulation of what the subject knows, whereas the subjective test, by allowing and encouraging the subject to express himself, evaluates the use he makes of his knowledge.

The word participation is a more accurate term than subjective and objective. It refers directly to the extent the subject takes part in the test, namely, the extent to which the subject speaks or writes in the language. An oral comprehension test can be either participating or non-participating. In the participating form the subject's comprehension is verified through his free response to a series of questions. In the non-participating form the subject's comprehension is verified through his selection of appropriate answers from a list of given answers after each question. The two categories, participating and non-participating, are not watertight, and from each extreme extends a whole gamut of sub-categories indicating increasing participation by the subject. The following examples with reading and oral comprehension tests will illustrate this.

Reading comprehension

1- Zero participation

The student reads a text. He reads a certain number of questions and selects what he believes to be the most appropriate one from a certain number of possible answers.

2- Limited participation

The student reads a text. He reads a certain number of questions and he is asked to limit his answer to a certain number of words.

3- Extended participation

The student reads a text. He is asked to make a summary.

4- Complete participation

The student reads a text. He is asked to comment on the text.

## Oral comprehension

1- Zero participation

The student hears a group of sentences. He is asked a question. He selects what he believes to be the most appropriate one from a certain given number of answers.

2- Limited participation

The student hears a group of sentences. He is asked a question. He gives his own answer.

3- Extended participation

The student hears a group of sentences. He is asked to make a summary.

4- Complete participation

The student hears a group of sentences. He is asked to comment.

The essential problem for the test designer is not so much how to evaluate what the subject says or writes but how to create a situation which will encourage the subject to give the maximum expression of which he is capable. The subject's response to the test situation is free and spontaneous; his production of language in the test is a representative sample of what he can produce in a variety of real situations.

Evaluation of the subject's production involves, under the respective headings of grammar, lexis, phonetics and socio-culture, four mains steps: identification; description; classification; written and numerical assessment as to the subject's language competence in a certain group of situations. The socio-cultural heading is optional and concerns the extent to which the subject is aware of the way of life of the community with which the language is associated. Linguistic theories apart, the evaluation procedure has to be simple, consistent and commu-

nicable; it should reflect the reality of the act of communication, speaking and writing, rather than some abstract view of what this reality is; it should provide the corrector with an objective (in the proper sense of the word) statement of what the subject produced, giving precise examples to back up or reject the interviewer's subjective impressions of the subject's fluency, and grammatical correction, for example.

The corrector has the very great advantage of basing his evaluation on what the subject can actually do with the language. This is not so with the non-participating form where the corrector just totals up a number of ticks and where not a word is said or written by the subject. In the non-participating form the language used in the test is selected by the test designer, not by the subject. The subject's attention is focussed by the constraint of the test on particular items chosen by the test designer. Answers to problems posed by these items are suggested by the test designer and the subject selects the most appropriate answer for resolving the problem. There is no guarantee that his selection of correct answers in the test will predict his capacity to use correct forms in speech or in writing.

It cannot however be denied that the non -participating form presents certain administrative advantages. Large numbers of subjects within a short period of time at an economic rate can be processed by non-specialist correctors, often with the aid of the computer.

The undeniable administrative conveniences conceal complex composition problems. There are no simple, widely accepted, easily applicable criteria as regards: what and which language items to include in the test; the proportion to observe between the items; the order in which the items should appear; the diffi-

culty of the items; the number of questions; the number of words and sentences per questions; the form of the answer; the administration of the test; the validation of the test. In the case of a grammar test, it will have to be decided whether the definite article will be tested or not. If it is to be tested, the test designer will have to determine how many times it will be tested and what particular grammatical function. When the grammatical items for testing have been selected, the test designer will then have to determine the difficulty of each item. It is no easy matter to determine exactly what it is that makes a question difficult, what it is that makes a question more difficult than another, and how much more difficult a question is than another. The task of the test designer is rendered all the more complex by the fact that the decisions he has to make are not purely based on linguistic factors, difficult though they may be to identify and to weigh, but on the continually changing needs and character of a particular group.

This is the great drawback of the non-participating test: it is designed with a certain group of subjects in mind. Consequently, it cannot be used for any other group. As the character of the group, as its need for the second or foreign language, and as teaching programmes change, the test will be in a constant state of revision. Should the test be administered throughout a region, a country or even beyond a country's borders, security problems oblige the test designer to have more than one version of any original. One is never sure whether the alternative version achieves the same results as the original.

Participating tests are free from these complex composition problems. The test designer of an oral expression test is only concerned with the appropriate means by which the interviewer can best encourage the subject to speak freely; and

thereby have sufficient evidence as to the subject's general capacity to speak. Unlike the non-participating oral comprehension test which only indicates how much the subject knows of the system of the language, the participating oral expression test indicates how much of the subject's knowledge of the language system is organised for the purpose of expression. Many students finish a course in the spoken language with high oral comprehension and poor oral expression: they speak very slowly and pause after every three to five words. This is because the knowledge they have acquired of the language during the course is passive, organised for oral comprehension. They have a good idea of how the language system is constructed; and they have learnt this through pattern drills. But as they have not used the structures of their pattern drills in real conversation concerning some motivating topic, they do not know how to speak; they can only understand.

The foregoing does not mean that non-participating tests serve little purpose. They have a very important rôle to play for inventory and diagnosis purposes: as the word inventory suggests, the test sets out to make a complete tabulation of what the subject knows of the way the language is organised; in the case of diagnosis, the test attempts to reject or confirm hypotheses as regards errors made by the subject. The fact that the subject may omit the auxiliary in the structure, "he <u>is</u> coming", may be due to phonetic or grammatical factors in an oral expression test. He may have trouble in saying the contracted form; he may or may not be aware of his mistake; he may even believe that he has pronounced the contracted form. Whatever the situation may be, the diagnosis test aims at explaining the mistake, and from that explanation appropriate corrective measures can be taken.

It is important not to use the non-participating test for other purposes than those of inventory and diagnosis. The only way to know if a person can speak or

write is to have him speak or write. It is even perhaps not exaggerated to say that the best way to test a person's reading or oral comprehension is to see whether he can talk or write about what he has read or heard. Does a person really understand a text if he cannot discuss it?

Evaluation of oral expression initially involves motivation and reality. The test situation must be motivating and real for the subject so that he speaks freely and so that his production is representative of his general capacity to speak in most situations. It does not matter whether written, spoken or graphic means are used to stimulate the subject as long as the means selected provide something motivating to talk about. Pictures are often objected to on the grounds that their socio-cultural content may prove a barrier to understanding. Such an objection is based on the erroneous assumption that language exists independently of the community which speaks it. In fact language is the expression of that community's way of life; and the teacher is obliged to teach both language and culture at the same time. Whether to teach British or American English is more than a question of which accent, but, and this is much more important, which way of life. The difference between the two cultures is well illustrated by these two sentences: "Can I buy you a drink?" and "Would you like a drink?" Teaching English to Amerindians is teaching them either the culture of the white man or an alternative language to use on the reservation. Be that as it may, in an oral expression test the subject is not being evaluated on his understanding of the pictures. The very fact that he may not understand may provide the necessary motivation and matter for conversation.

Motivation is directly related to the reality of the test situation. This reality has two aspects: the extent to which the test situation is real for the

subject; the extent to which the language generated in the test situation is representative of language generated in other situations. Certain situations are unsuitable for the test as little active use of the language is required to function efficiently in those situations: to have the subject buy an airline ticket or order food in a restaurant would not encourage the subject to speak a great deal, since what is required is minimal oral and reading comprehension, a few gestures and a few words.

The test should be divided into two sections, first interview and then discussion. The interview section, besides the obvious goal of putting the subject at his ease, is made up of personal questions which run through the whole gamut of what, when, where, who, why, how, be, can, do, have, will, would interrogations, which appear in most oral expression situations and are basic to the simplest conversation. A subject, who has difficulty in answering such fundamental questions as, "What is your name?", "How old are you?" and "How long have you been living in Québec?" cannot be expected to communicate effectively and easily in the most elementary of situations. The next section, discussion, goes beyond simple question and answer to establish to what extent the subject can talk about something that interests him. For an 8 year old it would be a question of seeing whether he could tell a simple story; for a foreign applicant to an English speaking university, whether he could discuss student participation. Such discussion, especially when time is limited and only a small number of interviewers are available, can be equally and even more effectively carried out at the level of the group. However, direction of a group, in the appropriate informal atmosphere, ensuring equal participation of subjects of different language competence, poses special problems. Both sections, interview and discussion, are necessary, as oral expression, whether it be at a party or a seminar, involves both question/

answer and narration. Many students never get beyond the question/answer stage and are incapable of the simplest account of something that has happened to them.

At the conclusion of the test, the interviewer's subjective impressions of the subject's performance should not be neglected, simply because they are intuitive and are not expressed in scientific language. Similarly, the opinions of teachers and others, after having listened to a certain number of recordings of interviews, should not be ignored. The subsequent analysis of the subject's oral production will provide the necessary concrete evidence, expressed in the appropriate, orderly, systematic scientific manner, to back up, modify or reject the interviewer's impressionistic judgement.

There are no widely accepted linguistic criteria of grammatical, lexical and phonetic correction, but there are two eminently practical criteria, which should underlie any evaluation, namely, comprehensibility and acceptability. Does the subject's error or deviation from the implicit and explicit norms of speech of a community make him difficult to understand, and if not, is that error or deviation acceptable to that community?

The absence of widely accepted linguistic criteria for the evaluation of oral expression does not mean that it is impossible, after sufficient samples of oral production have been taken, to set up criteria which can be easily translated into mechanical, formal and rapid identification, description and classification of what the subject says. The following grammatical, lexical and phonetic items, taken from actual tests at Laval as part of the English second language programme, will illustrate some of the evaluation problems:

Evaluation of errors

Grammatical and lexical errors

1- He takes breakfast at 9 o'clock.
2- He takes the breakfast at 9 o'clock.

3- He take breakfast at 9 o'clock.
4- He is taking the breakfast at 9 o'clock.
5- He taking the breakfast at 9 o'clock.
6- He participates to all their games.

Phonetic errors

Individual sounds

1- Bought is equivalent to bout.
2- Here's " he's.
3- Veal " bill.
4- Work " walk.

Stress and intonation

1- He's coming.

Risihg tone, instead of falling tone, on the word "coming" changes the statement into a question.

2- I don't like him.

Shift of primary stress from "like" to "him" involves two basic implicit meanings:

i) but I like her
ii) I hate him.

Evaluation of structures

Grammatical, lexical and phonetic structures

1- It's Bill.
2- He's here.
3- He's tired.
4- He's in the garden
5- He's coming.
6- He can come.
7- He's coming tomorrow at 5 o'clock.
8- He can come tomorrow at 5 o'clock.
9- He says (that) he can come tomorrow at 5 o'clock.
10- Although he's tired, he can come tomorrow at 5 o'clock.
11- He's too tired to come tomorrow at 5 o'clock.

Identification, description, classification and interpretation of such data is no more complex than writing a non-participating test and evaluating the results.

The above data seems to pose immense problems of assessment, but what is involved is

not a detailed phonetic analysis, worthy of a phonetician, but a rapid approximate survey. It is not difficult for a group of teachers and company personnel, after having listened to a tape, to arrive at an accurate assessment of the subject's stress, intonation and rhythm pattern, i.e., he puts the stress with a rising intonation on the last syllable of every third to fifth word, giving equal value to each syllable, pausing after every third to fifth word for three to five seconds, averaging some 90 words a minute. The general impression of the subject's performance in English is that he is difficult to understand, as he speaks slowly, as he pauses too often and too long, and his speech pattern is too centred on his mother tongue.

Such assessment is not in fact excessively complex, and the most significant feature of the assessment is that it is carried out on language produced by the subject and that his oral production can be directly compared with that of a mother tongue speaker. In a bilingual country where two languages enjoy official status and are both used as official languages in the various government agencies, the oral second language competence of civil servants can be evaluated in terms of the mother tongue speaker in specific job situations. Within the school system, the teacher and the programme director have a most effective way of assessing how successful the programme has been in teaching the spoken language and of comparing the student with a mother tongue speaker of the same age and similar background.

Whereas the assessment of the subject's oral production within the terms of the programme or in terms of the mother tongue speaker is based on criteria that change as the programme changes, and which vary according to the type of mother tongue speaker chosen as a model, the analysis of the oral production

should be standardised. Such standardisation permits the drawing up of a permanent record card of the student, which any director or any teacher, in any region, could interpret in terms of the arbitrary norms of his own programme. The record should be set out as follows:

- duration of test
- number of words produced by the subject
- number of sentences
- type of sentences
- type of grammatical structures
- type of lexical structures
- type of phonetic structures
- type of grammatical errors
- type of lexical errors
- type of phonetic errors

While in practice and for a long time to come, such standardised analysis will vary considerably from region to region, the tape recordings accompanying the student's record card will provide teachers with all the necessary information.

It will again be objected that such a standardised analysis under the above headings poses immense problems, as linguists are enormously divided on what language is, its function and how to describe and classify it. Such problems are no greater than those that must be faced in writing a non-participating test for a particular group, the results of which are as difficult to interpret as those of a participating test. Whatever the theories of linguists may be, for the teacher and the test designer language is a system, a system of systems, a dependent and independent structure, an individual act and a social phenomenon, an act of communication, a written and spoken code. What the teacher and the test designer require is a practical system of analysis, which corresponds to the reality of speech and writing.

The oral expression test used by CREDIF in Voix et Images de France as a classification and overall progress test, despite its limitations, indicates

how possible it is to arrive at a quick and accurate analysis of oral production. It is limited in its analysis: five minutes is too short a period of time to provide enough oral expression on which to base an analysis; there is no dialogue, only narration; the narration is confined to describing the action in a series of pictures centred on Parisian family life, the social class being somewhere between lower and middle bourgeoisie; the analysis of structures and errors is too complicated and too imprecise; the scoring of the errors, the coefficient of correction and expression, the interpretation of statistics obtained are only meaningful in terms of the experimental classes at CREDIF; no account is taken of the number of words produced by the subject, only the number of sentences.

Too much money and too much research has been concentrated on the non-participating (objective) test. Research should be concerned with analysing and assessing actual language production, namely, speaking and writing, and with the most effective test situation to encourage such production. As has already been pointed out, the non-participating test evaluates the subject's passive knowledge of the language, knowledge of the language's system for the purpose of comprehension. It gives no reliable indication of his active knowledge, how much that knowledge is geared for expression, for the purpose of communication. Is a university admission battery of non-participating tests in grammar, vocabulary and reading a real indication of the foreign student's capacity to use English?

It is often claimed that there is a high correlation between the performance of foreign students in these admission tests and their subsequent performance in their degree programme. However, success in their studies in not simply and necessarily a consequence of their second language competence in English, but the result of a whole series of variables, such as dedication and ability in their chosen discipline. Must not be forgotten those who passed the test and who made

a poor showing in their university courses. Of course the percentage of failures is small, but was their lack of success due to the test being unable to measure their second language competence or was it the result of other factors, such as, involvement in extra-curricular activities or lack of ability in their chosen discipline? Even assuming for the sake of argument that those who passed the test did well in their courses uniquely because of the language factor, it was not that the test predicted their language competence, but simply because, at that time of test, they were already proficient in speaking and writing. At its best, the battery of non-participating tests only indicates listening and reading fluency. Listening and reading fluency are little indication of speaking and writing fluency. University education is more than the capacity to understand the written or spoken word; it is the capacity to communicate, to discuss in speech or in writing what has been heard or read. It is this capacity to communicate that university admission tests in English for foreign students should be more concerned with.

The non-participating test's margin of error in rejecting those foreign students who do have the necessary second language competence in English should not be ignored, simply because the percentage involved is small. It is important to face the injustice done to the 20% who do not pass the test, but who do have the language competence required, and the hardship inflicted on the 20% who pass the test, but who do not have the required language competence.

The score which decides admission or rejection is difficult to determine and difficult to justify. Why should 80% rather than 70% be the pass mark? If a student falls short of the pass mark by 2, how important is that difference of 2 and what does it mean? Familiarity with non-participating tests can be the deciding factor for borderline students.

The administrative convenience of non-participating tests conceals their limitations. The importance of these limitations should by far outweigh the time and expense of participating tests, as these tests do measure what the subject can do with the language and how much his knowledge of the language has been and is organised for purposes of communication. Non-participating tests are often misunderstood and are often misused to the extent, sometimes, that the course is designed for the tests, rather than the tests for the course. In practice this may result in the teacher only teaching what can be measured by the non-participating tests. Thus a course in the spoken language may never get beyond listening fluency, the student's use of the language being limited to pattern drills. The knowledge the student acquires of the language is passive, organised for the purposes of oral comprehension, as he has little opportunity to organise what he learns for the purpose of communication in the real act of dialogue, trying to communicate an idea to another person.

In bilingual countries important decisions in government and business can be held up, because the second language competence of senior executives in oral comprehension is not matched by an equivalent second language competence in oral expression.

Whether the second or foreign language programme is taught to people at school or at university, to government or company employees, the student in the classroom has to have practice in communication. The most effective way, if not the only way, for the teacher, the programme director and other personnel to see how much a student has got out of his oral language programme is to place him in a situation where he has to talk to someone, to communicate, to exchange ideas in conversation. The role of the oral expression test is to do just that: create a situation in which the student participates freely and at his best.

# B I B L I O G R A P H Y


CREDIF (Centre de recherche et d'étude pour la diffusion du français) — Test C.G.M. 62. Paris: Didier, 1964, 52 p.

Alan Davies (Ed.) — Language testing symposium: a psycho-linguistic approach. London: Oxford University Press, 1968, 214 p.

Robert Lado — Language testing. London: Longmans, 1961, 389 p.

H.S. Otter — A functional language examination: the Modern Language Association Examination Project. London: Oxford University Press, 1968, 136 p.

George Perren — Testing ability in English as a second language. Part I. Problems. English Language Teaching, 21, 2(1967), p. 99-106.

Testing ability in English as a second language. Part 2. Techniques. English Language Teaching, 21, 3(1967), p. 197-202.

Testing ability in English as a second language. Part 3. Spoken language. English Language Teaching, 22, I (1967), p. 22-29.

G.D. Pickett — A comparison of translation and blank-filling as testing techniques. English Language Teaching, 23, 1(1968), p. 21-26.

John A. Upshur & Julia Fata (Eds.) — Problems in foreign language testing. Proceedings of a conference held at the University of Michigan, September 1967. Language Learning, Special Issue, No. 3, August 1968.

Rebecca Valette — Modern language testing: a handbook. New York: Harcourt, Brace & World, 1967, 200p.